

MEASURE TRANSFORMED QUASI SCORE TEST WITH APPLICATION TO LOCATION MISMATCH DETECTION

Koby Todros

Ben-Gurion University of the Negev

ABSTRACT

In this paper, we develop a generalization of the Gaussian quasi score test (GQST) for composite binary hypothesis testing. The proposed test, called measure transformed GQST (MT-GQST), is based on the score-function of the measure transformed Gaussian quasi maximum likelihood estimator (MT-GQMLE) that operates by empirically fitting a Gaussian model to a transformed probability measure of the data. By judicious choice of the transform we show that, unlike the GQST, the proposed MT-GQST involves higher-order statistical moments and can gain resilience to outliers, leading to significant mitigation of the model mismatch effect on the decision performance. A data-driven procedure for optimal selection of the measure transformation parameters is developed that minimizes the spectral norm of the empirical asymptotic error-covariance of the MT-GQMLE. This amounts to maximization of an empirical worst-case asymptotic local power at a fixed asymptotic size. The MT-GQST is applied to location mismatch detection of a near-field point source in a simulation example that illustrates its robustness to outliers.

Index Terms— Composite hypothesis testing, higher-order statistics, probability measure transform, robust statistics.

1. INTRODUCTION

The score test [1]–[3], also known as Rao’s score test, or the Lagrangian multiplier test, is a well established technique for composite binary hypothesis testing [1], [4], [5] that is based on the score-function¹ of the maximum likelihood estimator (MLE). Unlike the generalized likelihood ratio test (GLRT) [1], [5] and Wald’s test [1], [5], [6], it does not necessitate the maximum likelihood estimate under the alternative hypothesis, and therefore, may be significantly easier to compute. However, similarly to Wald’s test and the GLRT, it assumes knowledge of the likelihood function. In many practical scenarios the likelihood function is unknown, and therefore, alternatives to the score test become attractive.

A popular alternative of this kind is the Gaussian quasi score test (GQST) [7]–[11] that is based on the score-function of the Gaussian quasi MLE (GQMLE) [7], [11]–[16], which assumes that the samples obey Gaussian distribution. The GQST has gained popularity due to its implementation simplicity and ease of performance analysis that arise from the convenient Gaussian distribution. Despite the model mismatch, introduced by the normality assumption, the GQST has the appealing property of consistency under some mild regularity conditions [8]. However, in some circumstances, such as for certain types of non-Gaussian data, deviation from normality can inflict poor decision performance. This can occur when the first and second-order statistical moments are weakly identifiable over

the parameter space, or in the case of heavy-tailed data when the non-robust sample mean and covariance provide poor estimates in the presence of outliers.

In this paper, a generalization of the GQST is developed. The proposed generalization, called measure transformed GQST (MT-GQST), is based on the score-function of the measure transformed GQMLE (MT-GQMLE) [17], [18] that operates by empirically fitting a Gaussian model to a transformed probability measure of the data. The considered measure-transformation, also applied in [19]–[24], is structured by a non-negative function, called the MT-function, that weights the probability distribution of the data. By judicious choice of the MT-function we show that, unlike the GQST, the proposed MT-GQST involves higher-order statistical moments, can gain resilience to outliers, and yet have the computational and implementation advantages of the GQST.

Under some mild regularity conditions, we show that the MT-GQST is consistent. We also show that the asymptotic distribution of the test-statistic is central chi-squared under the null hypothesis, and non-central chi-squared under a sequence of local alternatives, with non-centrality parameter that is increasing with the inverse asymptotic error-covariance of the MT-GQMLE. A data driven procedure for optimal selection of the MT-function within some parametric class is developed that minimizes the spectral norm of the empirical asymptotic error-covariance of the MT-GQMLE. We show that this minimization amounts to maximization of an empirical worst-case asymptotic local power at a fixed asymptotic size.

The MT-GQST is illustrated for detecting a mismatch in the location of a near-field point source in the presence of spherically contoured noise. By specifying the MT-function within the family of zero-centred Gaussian functions parameterized by a scale parameter, we show that the MT-GQST outperforms the non-robust GQST and another robust alternative, and attains detection performance that are significantly closer to those of the omniscient score test that, unlike the MT-GQST, requires the knowledge of the likelihood function.

The paper is organized as follows. In Sections 2, the MT-GQMLE is reviewed. In Section 3, the score-function of the MT-GQMLE is used to construct the proposed MT-GQST. The proposed test is applied to location mismatch detection in Section 4. In Section 5, the main points of this contribution are summarized.

2. MEASURE TRANSFORMED GQMLE: REVIEW

We begin by reviewing the principles of the parametric probability measure transform [17], [18]. We then define a parametric measure-transformed mean vector and covariance matrix and show their relation to higher-order statistical moments. Furthermore, we formulate their strongly consistent estimates and state conditions for outlier resilience. Finally, these quantities are used to construct the MT-GQMLE [17], [18], whose objective function will be used in the following section to obtain the proposed MT-GQST.

¹Score-function of an estimator is referred here to as the gradient of the estimator’s objective function w.r.t. the vector parameter.

2.1. Probability measure transform

We define the measure space $(\mathcal{X}, \mathcal{S}_{\mathcal{X}}, P_{\mathbf{X};\theta})$, where $\mathcal{X} \subseteq \mathbb{C}^p$ is the observation space of a random vector \mathbf{X} , $\mathcal{S}_{\mathcal{X}}$ is a σ -algebra over \mathcal{X} and $P_{\mathbf{X};\theta}$ is an unknown probability measure on $\mathcal{S}_{\mathcal{X}}$ parameterized by a vector parameter θ that belongs to a parameter space $\Theta \subseteq \mathbb{R}^m$.

Definition 1. Given a non-negative function $u : \mathbb{C}^p \rightarrow \mathbb{R}_+$ satisfying

$$0 < \mathbb{E}[u(\mathbf{X}); P_{\mathbf{X};\theta}] < \infty, \quad (1)$$

where $\mathbb{E}[u(\mathbf{X}); P_{\mathbf{X};\theta}] \triangleq \int_{\mathcal{X}} u(\mathbf{x}) dP_{\mathbf{X};\theta}(\mathbf{x})$ and $\mathbf{x} \in \mathcal{X}$, a transform on $P_{\mathbf{X};\theta}$ is defined via the relation:

$$Q_{\mathbf{X};\theta}^{(u)}(A) \triangleq \mathbb{T}_u[P_{\mathbf{X};\theta}](A) = \int_A \varphi_u(\mathbf{x}; \theta) dP_{\mathbf{X};\theta}(\mathbf{x}), \quad (2)$$

where $A \in \mathcal{S}_{\mathcal{X}}$ and $\varphi_u(\mathbf{x}; \theta) \triangleq u(\mathbf{x})/\mathbb{E}[u(\mathbf{X}); P_{\mathbf{X};\theta}]$. The function $u(\cdot)$ is called the MT-function. [Proof: see Appendix A in [21]]

Proposition 1 (Properties of the transform). Let $Q_{\mathbf{X};\theta}^{(u)}$ be defined by relation (2). Then 1) $Q_{\mathbf{X};\theta}^{(u)}$ is a probability measure on $\mathcal{S}_{\mathcal{X}}$. 2) $Q_{\mathbf{X};\theta}^{(u)}$ is absolutely continuous w.r.t. $P_{\mathbf{X};\theta}$, with Radon-Nikodym derivative [25]:

$$dQ_{\mathbf{X};\theta}^{(u)}(\mathbf{x})/dP_{\mathbf{X};\theta}(\mathbf{x}) = \varphi_u(\mathbf{x}; \theta). \quad (3)$$

The probability measure $Q_{\mathbf{X};\theta}^{(u)}$ is said to be generated by the MT-function $u(\cdot)$.

2.2. The MT-mean and MT-covariance

According to (3) the mean vector and covariance matrix of \mathbf{X} under $Q_{\mathbf{X};\theta}^{(u)}$, that are assumed to be known parameterized functions of θ , are given by:

$$\boldsymbol{\mu}_{\mathbf{X};\theta}^{(u)} \triangleq \mathbb{E}[\mathbf{X}\varphi_u(\mathbf{X}; \theta); P_{\mathbf{X};\theta}] \quad (4)$$

and

$$\boldsymbol{\Sigma}_{\mathbf{X};\theta}^{(u)} \triangleq \mathbb{E}[\mathbf{X}\mathbf{X}^H \varphi_u(\mathbf{X}; \theta); P_{\mathbf{X};\theta}] - \boldsymbol{\mu}_{\mathbf{X};\theta}^{(u)} \boldsymbol{\mu}_{\mathbf{X};\theta}^{(u)H}, \quad (5)$$

respectively. Equations (4) and (5) imply that $\boldsymbol{\mu}_{\mathbf{X};\theta}^{(u)}$ and $\boldsymbol{\Sigma}_{\mathbf{X};\theta}^{(u)}$ are weighted mean and covariance of \mathbf{X} under $P_{\mathbf{X};\theta}$, with the weighting function $\varphi_u(\cdot; \cdot)$ defined below (2). By modifying the MT-function $u(\cdot)$, such that the condition (1) is satisfied, the MT-mean and MT-covariance under $Q_{\mathbf{X};\theta}^{(u)}$ are modified. In particular, by choosing $u(\cdot)$ to be any non-zero constant valued function we have $Q_{\mathbf{X};\theta}^{(u)} = P_{\mathbf{X};\theta}$, for which the standard mean vector $\boldsymbol{\mu}_{\mathbf{X};\theta}$ and covariance matrix $\boldsymbol{\Sigma}_{\mathbf{X};\theta}$ are obtained. Alternatively, when $u(\cdot)$ is non-constant analytic function, which has a convergent Taylor series expansion, the resulting MT-mean and MT-covariance involve higher-order statistical moments of $P_{\mathbf{X};\theta}$.

2.3. The empirical MT-mean and MT-covariance

Given a sequence of N i.i.d. samples from $P_{\mathbf{X};\theta}$ the empirical estimators of $\boldsymbol{\mu}_{\mathbf{X};\theta}^{(u)}$ and $\boldsymbol{\Sigma}_{\mathbf{X};\theta}^{(u)}$ are defined as:

$$\hat{\boldsymbol{\mu}}_{\mathbf{X}}^{(u)} \triangleq \sum_{n=1}^N \mathbf{X}_n \hat{\varphi}_u(\mathbf{X}_n) \quad (6)$$

and

$$\hat{\boldsymbol{\Sigma}}_{\mathbf{X}}^{(u)} \triangleq \sum_{n=1}^N \mathbf{X}_n \mathbf{X}_n^H \hat{\varphi}_u(\mathbf{X}_n) - \hat{\boldsymbol{\mu}}_{\mathbf{X}}^{(u)} \hat{\boldsymbol{\mu}}_{\mathbf{X}}^{(u)H}, \quad (7)$$

respectively, where $\hat{\varphi}_u(\mathbf{X}_n) \triangleq u(\mathbf{X}_n)/\sum_{n=1}^N u(\mathbf{X}_n)$. According to Proposition 2 in [21], if $\mathbb{E}[\|\mathbf{X}\|^2 u(\mathbf{X}); P_{\mathbf{X};\theta}] < \infty$ then

$\hat{\boldsymbol{\mu}}_{\mathbf{X}}^{(u)} \xrightarrow[N \rightarrow \infty]{\text{w.p. 1}} \boldsymbol{\mu}_{\mathbf{X};\theta}^{(u)}$ and $\hat{\boldsymbol{\Sigma}}_{\mathbf{X}}^{(u)} \xrightarrow[N \rightarrow \infty]{\text{w.p. 1}} \boldsymbol{\Sigma}_{\mathbf{X};\theta}^{(u)}$, where “ $\xrightarrow[\text{w.p. 1}]{} \rightarrow$ ” denotes convergence with probability (w.p.) 1 [26].

Robustness of the empirical MT-covariance (7) to outliers was studied in [21] using its influence function [27] which describes the effect on the estimator of an infinitesimal contamination at some point $\mathbf{y} \in \mathbb{C}^p$. An estimator is said to be B-robust if its influence function is bounded [27]. Similarly to the proof of Proposition 3 in [21] it can be shown that if the MT-function $u(\mathbf{y})$ and the product $u(\mathbf{y})\|\mathbf{y}\|^2$ are bounded over \mathbb{C}^p then the influence functions of both (6) and (7) are bounded.

2.4. The MT-GQMLE

Given a sequence of samples from $P_{\mathbf{X};\theta}$, the MT-GQMLE [18] of θ minimizes the empirical Kulback-Leibler divergence [28] between the transformed probability measure $Q_{\mathbf{X};\theta}^{(u)}$ and a complex circular Gaussian probability distribution $\Phi_{\mathbf{X};\theta}^{(u)}$ [29], characterized by the MT-mean $\boldsymbol{\mu}_{\mathbf{X};\theta}^{(u)}$ and MT-covariance $\boldsymbol{\Sigma}_{\mathbf{X};\theta}^{(u)}$. We have shown that this minimization amounts to maximization of the objective function:

$$J_u(\theta) \triangleq -D[\hat{\boldsymbol{\Sigma}}_{\mathbf{X}}^{(u)} \|\boldsymbol{\Sigma}_{\mathbf{X};\theta}^{(u)}] - \|\hat{\boldsymbol{\mu}}_{\mathbf{X}}^{(u)} - \boldsymbol{\mu}_{\mathbf{X};\theta}^{(u)}\|_{\boldsymbol{\Sigma}_{\mathbf{X};\theta}^{(u)}}^2, \quad (8)$$

where $D[\mathbf{A} \|\mathbf{B}] \triangleq \text{tr}[\mathbf{A}\mathbf{B}^{-1}] - \log \det[\mathbf{A}\mathbf{B}^{-1}] - p$ is the log-determinant divergence [30] between positive definite matrices \mathbf{A}, \mathbf{B} , $\|\mathbf{a}\|_{\mathbf{C}} \triangleq \sqrt{\mathbf{a}^H \mathbf{C} \mathbf{a}}$ denotes the weighted Euclidian norm of a vector \mathbf{a} with positive-definite weighting matrix \mathbf{C} and $\boldsymbol{\Sigma}_{\mathbf{X};\theta}^{(u)} \triangleq (\boldsymbol{\Sigma}_{\mathbf{X};\theta}^{(u)})^{-1}$. The MT-GQMLE is given by:

$$\hat{\theta}_u = \arg \max_{\theta \in \Theta} J_u(\theta). \quad (9)$$

Under some mild regularity conditions, we have shown that the MT-GQMLE is asymptotically normal and unbiased with convergence rate of $1/\sqrt{N}$, i.e., $\sqrt{N}(\hat{\theta}_u - \theta) \xrightarrow{D} \mathcal{N}(\mathbf{0}, \mathbf{R}_u(\theta))$ as $N \rightarrow \infty$, where “ \xrightarrow{D} ” denotes convergence in distribution [26]. The asymptotic error-covariance is given by

$$\mathbf{R}_u(\theta) = \mathbf{F}_u^{-1}(\theta) \mathbf{G}_u(\theta) \mathbf{F}_u^{-1}(\theta), \quad (10)$$

where

$$\mathbf{G}_u(\theta) \triangleq \mathbb{E}[u^2(\mathbf{X}) \boldsymbol{\psi}_u(\mathbf{X}; \theta) \boldsymbol{\psi}_u^T(\mathbf{X}; \theta); P_{\mathbf{X};\theta}], \quad (11)$$

$$\boldsymbol{\psi}_u(\mathbf{X}; \theta) \triangleq \nabla_{\theta} \log \phi^{(u)}(\mathbf{X}; \theta), \quad (12)$$

$$\mathbf{F}_u(\theta) \triangleq -\mathbb{E}[u(\mathbf{X}) \boldsymbol{\Gamma}_u(\mathbf{X}; \theta); P_{\mathbf{X};\theta}], \quad (13)$$

$$\boldsymbol{\Gamma}_u(\mathbf{X}; \theta) \triangleq \nabla_{\theta}^2 \log \phi^{(u)}(\mathbf{X}; \theta), \quad (14)$$

$\phi^{(u)}(\mathbf{X}; \theta)$ is the density of the Gaussian measure $\Phi_{\mathbf{X};\theta}^{(u)}$ and it is assumed that $\mathbf{F}_u(\theta)$ is non-singular.

3. THE MEASURE TRANSFORMED GAUSSIAN QUASI SCORE TEST

Given a sequence of samples from $P_{\mathbf{X};\theta}$, we use the score-function of the MT-GQMLE (9) to construct the proposed MT-GQST for the composite hypothesis testing problem:

$$H_0 : \theta = \theta_0 \quad (15)$$

$$H_1 : \theta \neq \theta_0.$$

Under some mild regularity conditions we show that the MT-GQST is consistent. Furthermore, we derive the asymptotic distribution of its test-statistic under the null hypothesis and under a sequence of local alternatives. Finally, a data driven procedure for optimal selection of the MT-function is developed.

3.1. The MT-GQST

Notice that by (6) and (7) the objective function (8) is an empirical estimate of $\bar{J}_u(\boldsymbol{\theta}, \boldsymbol{\vartheta}) \triangleq -D[\Sigma_{\mathbf{X};\boldsymbol{\vartheta}}^{(u)} \|\Sigma_{\mathbf{X};\boldsymbol{\vartheta}}^{(u)}] - \|\boldsymbol{\mu}_{\mathbf{X};\boldsymbol{\vartheta}}^{(u)} - \boldsymbol{\mu}_{\mathbf{X};\boldsymbol{\vartheta}}^{(u)}\|_{\Omega_{\mathbf{X};\boldsymbol{\vartheta}}^{(u)}}^2$.

One can verify that when $\bar{J}_u(\boldsymbol{\theta}, \boldsymbol{\vartheta})$ is $\boldsymbol{\vartheta}$ -differentiable it has a stationary point at $\boldsymbol{\vartheta} = \boldsymbol{\theta}$. Assuming that this stationary point is unique $\nabla_{\boldsymbol{\vartheta}} \bar{J}_u(\boldsymbol{\theta}, \boldsymbol{\theta}_0) = \mathbf{0}$ when $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ and $\nabla_{\boldsymbol{\vartheta}} \bar{J}_u(\boldsymbol{\theta}, \boldsymbol{\theta}_0) \neq \mathbf{0}$ when $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$. This motivates the use of $\nabla J_u(\boldsymbol{\theta})$, i.e., the score-function of the MT-GQMLE, for testing between H_0 and H_1 . Hence, we define the normalized score-function: $\boldsymbol{\eta}_u(\boldsymbol{\theta}) \triangleq \nabla J_u(\boldsymbol{\theta}) \times (1/\sqrt{N}) \sum_{n=1}^N u(\mathbf{X}_n)$. By (6)-(8)

$$\boldsymbol{\eta}_u(\boldsymbol{\theta}) = (1/\sqrt{N}) \sum_{n=1}^N u(\mathbf{X}_n) \boldsymbol{\psi}_u(\mathbf{X}_n; \boldsymbol{\theta}), \quad (16)$$

where $\boldsymbol{\psi}_u(\mathbf{X}; \boldsymbol{\theta})$ is defined in (12). Furthermore, we define the empirical estimate of (11):

$$\hat{\mathbf{G}}_u(\boldsymbol{\theta}) \triangleq N^{-1} \sum_{n=1}^N u^2(\mathbf{X}_n) \boldsymbol{\psi}_u(\mathbf{X}_n; \boldsymbol{\theta}) \boldsymbol{\psi}_u^T(\mathbf{X}_n; \boldsymbol{\theta}). \quad (17)$$

The MT-GQST for the hypothesis testing problem (15) is defined as:

$$T_u \triangleq \boldsymbol{\eta}_u^T(\boldsymbol{\theta}_0) \hat{\mathbf{G}}_u^{-1}(\boldsymbol{\theta}_0) \boldsymbol{\eta}_u(\boldsymbol{\theta}_0) \underset{H_0}{\overset{H_1}{\geq}} t, \quad (18)$$

where $t \in \mathbb{R}_+$ denotes a threshold. By modifying the MT-function $u(\cdot)$ such that condition (1) is satisfied the MT-GQST is modified, resulting in a family of tests. In particular, when $u(\cdot)$ is any non-zero constant function $Q_{\mathbf{X};\boldsymbol{\theta}}^{(u)} = P_{\mathbf{X};\boldsymbol{\theta}}$ and the standard GQST is obtained that only involves first and second-order statistical moments.

3.2. Asymptotic performance analysis

Here, we study the asymptotic performance of the proposed test (18). We assume that a sequence of i.i.d. samples \mathbf{X}_n , $n = 1, \dots, N$ from $P_{\mathbf{X};\boldsymbol{\theta}}$ is available and that the parameter space $\boldsymbol{\Theta}$ is compact. We begin by stating some regularity conditions that will be used in the sequel:

- (A-1) $E[u(\mathbf{X}) \boldsymbol{\psi}_u(\mathbf{X}; \boldsymbol{\theta}_0); P_{\mathbf{X};\boldsymbol{\theta}}] \neq \mathbf{0}$ for $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$.
- (A-2) $E[u^2(\mathbf{X}) \boldsymbol{\psi}_u(\mathbf{X}; \boldsymbol{\theta}_0) \boldsymbol{\psi}_u^T(\mathbf{X}; \boldsymbol{\theta}_0); P_{\mathbf{X};\boldsymbol{\theta}}]$ is non-singular.
- (A-3) $\boldsymbol{\mu}_{\mathbf{X};\boldsymbol{\theta}}^{(u)}$ and $\Sigma_{\mathbf{X};\boldsymbol{\theta}}^{(u)}$ are twice continuously differentiable in $\boldsymbol{\Theta}$.
- (A-4) $E[u^4(\mathbf{X}); P_{\mathbf{X};\boldsymbol{\theta}}]$ and $E[\|\mathbf{X}\|^8 u^4(\mathbf{X}); P_{\mathbf{X};\boldsymbol{\theta}}]$ are bounded.
- (A-5) $\mathbf{G}_u(\boldsymbol{\theta})$ is bounded and non-singular.
- (A-6) The density of $P_{\mathbf{X};\boldsymbol{\theta}}$ is continuous in $\boldsymbol{\Theta}$ a.e. over \mathcal{X} .
- (A-7) The Fisher information matrix $\mathbf{I}_{\text{FIM}}(\boldsymbol{\theta})$ [29] is bounded.

The following proposition states consistency conditions.

Proposition 2 (Consistency). *Assume that conditions A-1–A-4 are satisfied. Then, for any $t \in \mathbb{R}$*

$$\Pr[T_u > t] \xrightarrow[N \rightarrow \infty]{} 1 \quad \text{under } H_1. \quad (19)$$

[A proof is given in Appendix A]

Next, we derive the asymptotic distribution of the test-statistic under the null hypothesis and under a sequence of local alternatives.

Proposition 3 (Asymptotic distribution under the null hypothesis). *Assume that conditions A-3–A-5 are satisfied. Then,*

$$T_u \xrightarrow[N \rightarrow \infty]{D} \chi_m^2 \quad \text{under } H_0, \quad (20)$$

where χ_m^2 denotes a central chi-squared distribution with m -degrees of freedom. [A proof appears in Appendix B]

Theorem 1 (Asymptotic distribution under local alternatives). *Assume that conditions A-3–A-7 are satisfied. Furthermore, consider a sequence of local alternatives that converges to $\boldsymbol{\theta}_0$ at a rate of $1/\sqrt{N}$. Specifically, consider*

$$H_1 : \boldsymbol{\theta} = \boldsymbol{\theta}_0 + \mathbf{h}/\sqrt{N}, \quad (21)$$

where $\mathbf{h} \in \mathbb{R}^m$ is a non-zero locality parameter. Then,

$$T_u \xrightarrow[N \rightarrow \infty]{D} \chi_m^2(\lambda_u(\mathbf{h})) \quad \text{under } H_1, \quad (22)$$

where $\chi_m^2(\lambda_u(\mathbf{h}))$ is a non-central chi-squared distribution with m -degrees of freedom and non-centrality parameter $\lambda_u(\mathbf{h}) \triangleq \mathbf{h}^T \mathbf{R}_u^{-1}(\boldsymbol{\theta}_0) \mathbf{h}$. The matrix $\mathbf{R}_u(\cdot)$ is the asymptotic error-covariance (10) of the MT-GQMLE (9). [A proof appears in Appendix C]

The following Corollary is a direct consequence of (22), the Rayleigh-Ritz Theorem [41] and the property that the right-tail probability of the non-central chi-squared distribution is monotonically increasing in the non-centrality parameter [32].

Corollary 1 (Asymptotic local power). *Assume that conditions A-3–A-7 hold. Under the local alternatives (21), the asymptotic power at a fixed asymptotic size α satisfies*

$$\beta_u^{(\alpha)}(\mathbf{h}) = Q_{\chi_m^2(\lambda_u(\mathbf{h}))} \left(Q_{\chi_m^2}^{-1}(\alpha) \right), \quad (23)$$

where $Q_{\chi_m^2}(\cdot)$ and $Q_{\chi_m^2(\cdot)}(\cdot)$ denote the right-tail probabilities of the central and non-central chi-squared distributions, respectively. Furthermore, for any $c > 0$ the worst-case asymptotic power

$$\bar{\beta}_u^{(\alpha)}(c) \triangleq \min_{\mathbf{h}: \|\mathbf{h}\| \geq c} \beta_u^{(\alpha)}(\mathbf{h}) = Q_{\chi_m^2(\gamma_u(c))} \left(Q_{\chi_m^2}^{-1}(\alpha) \right), \quad (24)$$

where $\gamma_u(c) \triangleq c^2 \|\mathbf{R}_u(\boldsymbol{\theta}_0)\|_S^{-1}$ and $\|\cdot\|_S$ denotes the spectral norm.

3.3. Selection of the MT-function

While according to Propositions 2 and 3, the asymptotic global power and size are invariant to the choice of the MT-function $u(\cdot)$, by Corollary 1 one sees that it controls the asymptotic local power through the error-covariance $\mathbf{R}_u(\boldsymbol{\theta}_0)$ (10). Since the tail probability of the non-central chi-squared distribution is monotonically increasing in the non-centrality parameter [32], minimization of the spectral norm $\|\mathbf{R}_u(\boldsymbol{\theta}_0)\|_S$ amounts to maximization of the worst case asymptotic local power $\bar{\beta}_u^{(\alpha)}(c)$ (24) for any fixed c and asymptotic size α . Hence, we propose to choose $u(\cdot)$ that minimizes $\|\hat{\mathbf{R}}_u(\boldsymbol{\theta}_0)\|_S$, where $\hat{\mathbf{R}}_u(\boldsymbol{\theta})$ is an empirical estimate of error-covariance (10) defined as:

$$\hat{\mathbf{R}}_u(\boldsymbol{\theta}) \triangleq \hat{\mathbf{F}}_u^{-1}(\boldsymbol{\theta}) \hat{\mathbf{G}}_u(\boldsymbol{\theta}) \hat{\mathbf{F}}_u^{-1}(\boldsymbol{\theta}), \quad (25)$$

where $\hat{\mathbf{F}}_u(\boldsymbol{\theta}) \triangleq -N^{-1} \sum_{n=1}^N u(\mathbf{X}_n) \boldsymbol{\Gamma}_u(\mathbf{X}_n; \boldsymbol{\theta})$ is an estimate of (13) and $\hat{\mathbf{G}}_u(\boldsymbol{\theta})$ is defined in (17). It can be shown that if conditions A-3, A-4, A-5, A-7, are satisfied then $\hat{\mathbf{R}}_u(\boldsymbol{\theta}_0) \xrightarrow[N \rightarrow \infty]{P} \mathbf{R}_u(\boldsymbol{\theta}_0)$ under the local alternatives (21).

Here, we restrict the class of MT-functions to some parametric family $\{u(\mathbf{X}; \boldsymbol{\omega}), \boldsymbol{\omega} \in \Omega \subseteq \mathbb{C}^r\}$ that satisfies the conditions stated in Definition 1 and Theorem 1. For example, the Gaussian family of functions that satisfy the robustness conditions stated at the ending paragraph of Subsection 2.3 is a natural choice for inducing outlier resilience. Hence, an optimal choice of the MT-function parameter $\boldsymbol{\omega}$ would minimize $\|\hat{\mathbf{R}}_u(\boldsymbol{\theta}_0)\|_S$ that is constructed from (25) by the same sequence of samples used for obtaining the MT-GQST (18).

4. NUMERICAL EXAMPLE

We consider the problem of detecting a mismatch in the location of an emitting narrowband near-field point source that is formulated as the following composite binary hypothesis testing problem:

$$\begin{aligned} H_0 &: \mathbf{X}_n = S_n \mathbf{a}(\boldsymbol{\theta}_0) + \mathbf{W}_n, \quad n = 1, \dots, N, \\ H_1 &: \mathbf{X}_n = S_n \mathbf{a}(\boldsymbol{\theta}) + \mathbf{W}_n, \quad \boldsymbol{\theta} \neq \boldsymbol{\theta}_0, \quad n = 1, \dots, N, \end{aligned} \quad (26)$$

where $\{\mathbf{X}_n \in \mathbb{C}^p\}$ is an observation process, $\{S_n \in \mathbb{C}\}$ is an i.i.d. symmetrically distributed random signal process, and $\{\mathbf{W}_n \in \mathbb{C}^p\}$ is an i.i.d. noise process that is statistically independent of $\{S_n\}$. We assume that the noise component is spherically contoured [33] with stochastic representation $\mathbf{W}_n = \nu_n \mathbf{Z}_n$, where $\{\nu_n \in \mathbb{R}_{++}\}$ is an i.i.d. process and $\{\mathbf{Z}_n \in \mathbb{C}^p\}$ is a proper-complex wide-sense stationary Gaussian process with zero-mean and scaled unit covariance $\sigma_Z^2 \mathbf{I}$. The processes $\{\nu_n\}$ and $\{\mathbf{Z}_n\}$ are assumed to be statistically independent. The vector $\mathbf{a}(\boldsymbol{\theta})$, $\boldsymbol{\theta} \triangleq [r, \vartheta]^T$, is the steering vector of a uniform linear array of p sensors with inter-element spacing d that receive a signal with wavelength λ generated by a narrowband near-field point source with range r and bearing ϑ . By Fresnel's approximation [34], [35] when $0.62(d^3(p-1)^3/\lambda)^{1/2} < r < 2d^2(p-1)^2/\lambda$ we have $[\mathbf{a}(\boldsymbol{\theta})]_k = \exp(j(\omega_e k + \phi_e k^2 + O(d^2/r^2)))$, $k = 0, \dots, p-1$, where $\omega_e \triangleq -2\pi d \sin(\vartheta)/\lambda$ and $\phi_e \triangleq \pi d^2 \cos^2(\vartheta)/(\lambda r)$ are called “electrical angles”.

In order to gain robustness against outliers, as well as sensitivity to higher-order moments, we specify the MT-function in the zero-centred Gaussian family of functions parametrized by a width parameter ω , i.e.,

$$u(\mathbf{x}; \omega) = \exp(-\|\mathbf{x}\|^2/\omega^2), \quad \omega \in \mathbb{R}_{++}. \quad (27)$$

Notice that the MT-function (27) satisfies the robustness conditions stated at the ending paragraph of Subsection 2.3. To obtain the corresponding test-statistic (18) and the empirical error-covariance (25), one has to compute the vector and matrix functions $\psi_u(\cdot; \boldsymbol{\theta})$ and $\Gamma_u(\cdot; \boldsymbol{\theta})$, defined in (12) and (14), respectively. By (4), (5), (12), (14), (26) and (27) these quantities take the following simple forms:

$$\psi_u(\mathbf{X}; \boldsymbol{\theta}) = \xi(\omega) \nabla_{\boldsymbol{\theta}} |\mathbf{X}^H \mathbf{a}(\boldsymbol{\theta})|^2,$$

$$\Gamma_u(\mathbf{X}; \boldsymbol{\theta}) = \xi(\omega) \begin{bmatrix} \nabla_r^2 |\mathbf{X}^H \mathbf{a}(\boldsymbol{\theta})|^2 & \nabla_r \nabla_{\vartheta} |\mathbf{X}^H \mathbf{a}(\boldsymbol{\theta})|^2 \\ \nabla_r \nabla_{\vartheta} |\mathbf{X}^H \mathbf{a}(\boldsymbol{\theta})|^2 & \nabla_{\vartheta}^2 |\mathbf{X}^H \mathbf{a}(\boldsymbol{\theta})|^2 \end{bmatrix},$$

where $\xi(\omega)$ is a strictly positive functions of ω . It is important to note that the resulting test-statistic (18) and the empirical error-covariance (25) are independent of $\xi(\omega)$.

In the following simulation we evaluate the detection performance of the proposed MT-GQST as compared to the score test, the GQST, and another robust GQST extension, called here ZMNL-GQST. The ZMNL-GQST operates by applying GQST after passing the data through a zero-memory non-linear (ZMNL) function that suppresses outliers by clipping the amplitude of the observations.

We use the same ZMNL preprocessing approach that has been applied in [36]–[38] to robustify the MUSIC algorithm [16].

We considered a BPSK signal with variance σ_S^2 impinging on $p = 8$ sensors with inter-element spacing $d = \lambda/4 = 0.25$ [m]. Two types of noise distributions were examined: 1) Gaussian and 2) heavy-tailed K -distributed noise [33] with shape parameter $\kappa = 0.75$. The sample size was set to $N = 1000$. The signal-to-noise-ratio (SNR), used to index the detection performance, is defined as $\text{SNR} \triangleq 10 \log_{10} \sigma_S^2 / \sigma_Z^2$. The location vector parameter at the null was set to $\boldsymbol{\theta}_0 = [r_0, \vartheta_0]^T$, where $r_0 = 1.5$ [m] and $\vartheta_0 = 0^\circ$. We considered a specific local alternative $\boldsymbol{\theta}_1 = [r_1, \vartheta_1]^T$, corresponding to $\mathbf{h} = \sqrt{N}(\boldsymbol{\theta}_1 - \boldsymbol{\theta}_0)$ in (21), where $r_1 = r_0 + 0.01$ [m] and $\vartheta_1 = \vartheta_0 + 0.5^\circ$. The optimal width parameter ω_{opt} of Gaussian MT-function (27) was obtained by minimizing the spectral norm $\|\hat{\mathbf{R}}_u(\boldsymbol{\theta}_0)\|_S$ of the empirical error-covariance (25) over $\Omega = [1, 30]$. All empirical power curves were obtained using 10^4 Monte-Carlo simulations.

Fig. 1 depicts the empirical and asymptotic (23) power curves of the MT-GQST as compared to the empirical power curves of the GQST, ZMNL-GQST and the score test for a fixed test size equal to 10^{-2} . Notice that when the noise is Gaussian, the MT-GQST, GQST and ZMNL-GQST attain similar performance. For the K -distributed noise, the MT-GQST outperforms the GQST and ZMNL-GQST, and significantly reduces the gap towards the score test, which unlike the MT-GQST, assumes knowledge of the likelihood function.

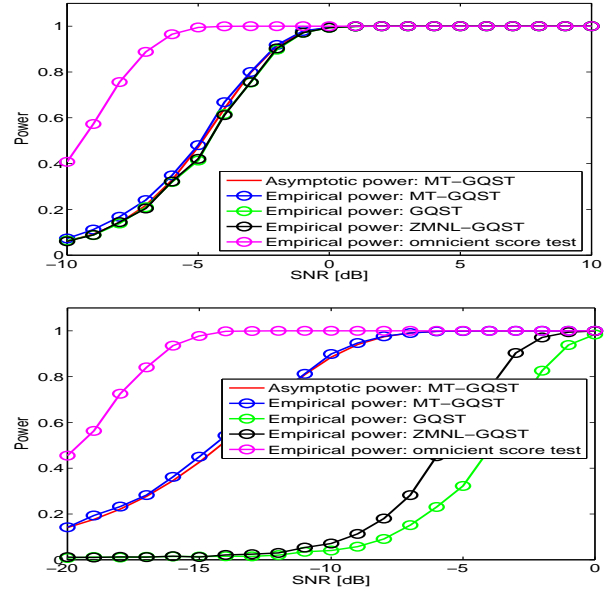


Fig. 1. Location mismatch detection in Gaussian noise (top) and K -distributed noise (bottom).

5. CONCLUSION

In this paper a new score-type test, called MT-GQST, was derived based on the score-function of the measure transformed GQMLE. By specifying the MT-function in the Gaussian family, the proposed test was applied to location mismatch detection in non-Gaussian noise. Exploration of other MT-functions may result in additional tests in this class that have different useful properties.

Appendices

A. Proof of Proposition 2:

We first show that $\eta_u(\theta_0)$ and $\hat{G}_u(\theta_0)$ satisfy

$$\frac{1}{\sqrt{N}}\eta_u(\theta_0) \xrightarrow[N \rightarrow \infty]{\text{w.p. 1}} E[u(\mathbf{X}) \psi_u(\mathbf{X}; \theta_0); P_{\mathbf{X}; \theta}] \quad (28)$$

and

$$\hat{G}_u(\theta_0) \xrightarrow[N \rightarrow \infty]{\text{w.p. 1}} E[u^2(\mathbf{X}) \psi_u(\mathbf{X}; \theta_0) \psi_u^T(\mathbf{X}; \theta_0); P_{\mathbf{X}; \theta}]. \quad (29)$$

Since $\{\mathbf{X}_n\}_{n=1}^N$ are i.i.d. random variables and the functions $u(\cdot)$ and $\psi_u(\cdot, \cdot)$ are real, the products $\{u(\mathbf{X}_n) \psi_u(\mathbf{X}_n, \theta_0)\}_{n=1}^N$ and $\{u(\mathbf{X}_n) \psi_u(\mathbf{X}_n, \theta_0) \psi_u^T(\mathbf{X}_n, \theta_0)\}_{n=1}^N$, comprising (16) and (17), respectively, are real and i.i.d. Furthermore, by (12), (60) and assumption A-3 it can be shown that there exists a positive constant B such that $|\psi_u(\mathbf{X}, \theta)|_k \leq B \sum_{l=0}^2 \|\mathbf{X}\|^l$, $k = 1, \dots, p$ and $|\psi_u(\mathbf{X}; \theta) \psi_u^T(\mathbf{X}; \theta)|_{k,j} \leq B \sum_{l=0}^4 \|\mathbf{X}\|^l$, $k, j = 1, \dots, p$. Hence, according to (1), Assumption A-4 and Hölder's inequality [26] the expectations $\{E[u(\mathbf{X}) |\psi_u(\mathbf{X}, \theta_0)]_k; P_{\mathbf{X}; \theta}\}_{k=1}^p$ and $\{E[u^2(\mathbf{X}) |\psi_u(\mathbf{X}; \theta_0) \psi_u^T(\mathbf{X}; \theta_0)]_{k,j}; P_{\mathbf{X}; \theta}\}_{k,j=1}^p$ are finite. Therefore, by (16), (17) and Khinchine's strong law of large numbers [25] relations (28) and (29) must hold.

Thus, by (18), (28), (29), Assumptions A-1, A-2 and the Mann-Wald theorem [39] we conclude that $\frac{1}{N}T_u \xrightarrow[N \rightarrow \infty]{\text{w.p. 1}} C$ under H_1 , where C denotes some positive constant. The relation (19) directly follows. \square

B. Proof of Proposition 3:

Under assumptions A-3 and A-4 it is shown in Lemma 5 stated in Appendix B in [18] that the normalized score function (16) satisfies

$$\eta_u(\theta_0) \xrightarrow[N \rightarrow \infty]{D} \mathcal{N}(\mathbf{0}, \mathbf{G}_u(\theta_0)). \quad (30)$$

Furthermore, by assumptions A-3 and A-4 it also follows from relation (29) in Appendix A that under H_0 ($P_{\mathbf{X}; \theta} = P_{\mathbf{X}; \theta_0}$)

$$\hat{G}_u(\theta_0) \xrightarrow[N \rightarrow \infty]{\text{w.p. 1}} \mathbf{G}_u(\theta_0), \quad (31)$$

where according to assumption A-5 $\mathbf{G}_u(\theta_0)$ is non-singular. Hence, relation (20) follows from (30), (31), Slutsky's Theorem [26], Mann-Wald's Theorem [39] and the properties of quadratic forms of Gaussian random variables [5]. \square

C. Proof of Theorem 1:

In Propositions 4 and 5 stated below we show that $\eta_u(\theta_0) \xrightarrow[N \rightarrow \infty]{D} \mathcal{N}(\mathbf{F}_u(\theta_0) \mathbf{h}, \mathbf{G}_u(\theta_0))$ and $\hat{G}_u(\theta_0) \xrightarrow[N \rightarrow \infty]{P} \mathbf{G}_u(\theta_0)$. Hence, by (18) relation (22) follows from Slutsky's Theorem [26], Mann-Wald's Theorem [39] and the properties of quadratic forms of Gaussian random variables [5]. \square

The following lemmas are based on the fact that since by (21) the parameter θ changes with the sample size N , the observations form a triangular array [40] (rather than a sequence) of random vectors:

$$\mathbf{X}_{N,k}, \quad k = 1, \dots, N, \quad N \geq 1, \quad (32)$$

where $\mathbf{X}_{N,k}$, $k = 1, \dots, N$ are i.i.d. with probability distribution $P_{\mathbf{X}; \theta_0 + \frac{\mathbf{h}}{\sqrt{N}}}$.

Proposition 4. Assume that conditions A-3-A-7 are satisfied. Under the local alternatives (21), the normalized score function $\eta_u(\theta_0)$ (16) satisfies:

$$\eta_u(\theta_0) \xrightarrow[N \rightarrow \infty]{D} \mathcal{N}(\mathbf{F}_u(\theta_0) \mathbf{h}, \mathbf{G}_u(\theta_0)). \quad (33)$$

Proof. By assumption A-3, the vector function $\psi_u(\mathbf{x}; \theta)$ defined in (12) is continuous in Θ for any $\mathbf{x} \in \mathcal{X}$. Therefore, by Identity 1, the normalized score function (16) satisfies $\eta_u(\theta_0) = \eta_u(\theta) + \hat{\mathbf{F}}_u(\theta^*) \mathbf{h}$, where $\hat{\mathbf{F}}_u(\cdot)$ is defined below (25) and θ^* lies in the line segment connecting θ and θ_0 . Furthermore, by Lemmas 1 and 2 stated below in Appendix D, $\eta_u(\theta) \xrightarrow[N \rightarrow \infty]{D} \mathcal{N}(\mathbf{0}, \mathbf{G}_u(\theta_0))$ and $\hat{\mathbf{F}}_u(\theta^*) \xrightarrow[N \rightarrow \infty]{P} \mathbf{F}_u(\theta_0)$. Therefore, the relation (33) follows directly from Slutsky's Theorem [26]. \square

Proposition 5. Assume that conditions A-3, A-4, A-6, A-7 are satisfied. Then, under (21)

$$\hat{G}_u(\theta_0) \xrightarrow[N \rightarrow \infty]{P} \mathbf{G}_u(\theta_0). \quad (34)$$

Proof. Define a triangular array of real random variables obtained from the array (32):

$$Y_{N,k} \triangleq g(\mathbf{X}_{N,k}), \quad k = 1, \dots, N, \quad N \geq 1, \quad (35)$$

where

$$g(\mathbf{X}) \triangleq u^2(\mathbf{X}) [\psi_u(\mathbf{X}; \theta_0)]_l [\psi_u(\mathbf{X}; \theta_0)]_m. \quad (36)$$

Since $\mathbf{X}_{N,k}$, $k = 1, \dots, N$ are i.i.d. and the functions $u(\cdot)$ and $\psi_u(\cdot; \cdot)$ are real, then $Y_{N,k}$, $k = 1, \dots, N$ are real and i.i.d. Furthermore, define the random variable

$$Y \triangleq g(\mathbf{X}), \quad (37)$$

where \mathbf{X} has probability distribution $P_{\mathbf{X}; \theta_0}$.

Let $F_{Y_{N,1}}(\cdot)$ and $F_Y(\cdot)$ denote the c.d.f.s of $Y_{N,1}$ and Y , respectively. We show that

$$F_{Y_{N,1}}(y) \xrightarrow[N \rightarrow \infty]{} F_Y(y) \quad \forall y \in C, \quad (38)$$

where $C \subseteq \mathbb{R}$ denotes the set of continuity points of $F_Y(y)$. Let $\zeta_{Y_{N,1}}(t)$ and $\zeta_Y(t)$ denote the characteristic functions of $Y_{N,1}$ and Y , respectively. By (35) and (37) their difference satisfies $|\zeta_{Y_{N,1}}(t) - \zeta_Y(t)| = \int_{\mathcal{X}} e^{itg(\mathbf{x})} (f(\mathbf{x}; \theta) - f(\mathbf{x}; \theta_0)) d\rho(\mathbf{x})$, where $f(\mathbf{x}; \theta) \triangleq dP_{\mathbf{X}; \theta}(\mathbf{x})/d\rho(\mathbf{x})$ is the density function of $P_{\mathbf{X}; \theta}$ w.r.t a dominating σ -finite measure ρ on \mathcal{X} . Hence, by (21), (65) and assumption A-6

$$\begin{aligned} |\zeta_{Y_{N,1}}(t) - \zeta_Y(t)| &\stackrel{(a)}{\leq} \frac{\int_{\mathcal{X}} |\mathbf{h}^T \boldsymbol{\eta}(\mathbf{X}; \theta^*)| f(\mathbf{x}; \theta^*) d\rho(\mathbf{x})}{\sqrt{N}} \\ &= \frac{E[|\mathbf{h}^T \boldsymbol{\eta}(\mathbf{X}; \theta^*)|; P_{\mathbf{X}; \theta^*}]}{\sqrt{N}} \\ &\stackrel{(b)}{\leq} \frac{\sqrt{E[|\mathbf{h}^T \boldsymbol{\eta}(\mathbf{X}; \theta^*)|^2; P_{\mathbf{X}; \theta^*}]}}{\sqrt{N}} \\ &= \frac{\|\mathbf{h}\|_{\text{FIM}(\theta^*)}}{\sqrt{N}}, \end{aligned} \quad (39)$$

where $\boldsymbol{\eta}(\mathbf{x}; \boldsymbol{\theta}) \triangleq \nabla_{\boldsymbol{\theta}} \log f(\mathbf{x}; \boldsymbol{\theta})$, $\mathbf{I}_{\text{FIM}} \triangleq \mathbb{E}[\boldsymbol{\eta}(\mathbf{x}; \boldsymbol{\theta}) \boldsymbol{\eta}^T(\mathbf{x}; \boldsymbol{\theta})]$ is the Fisher-information matrix [42], $\boldsymbol{\theta}^*$ lies in the line segment connecting $\boldsymbol{\theta} = \boldsymbol{\theta}_0 + \frac{\mathbf{h}}{\sqrt{N}}$ and $\boldsymbol{\theta}_0$, (a) follows from the triangle inequality, and (b) follows from Hölder's inequality [26]. Therefore, by (39) and (62) $\zeta_{Y_N}(t) \xrightarrow{N \rightarrow \infty} \zeta_Y(t)$ for any $t \in \mathbb{R}$. Hence, by Theorem 11.2.2 in [1] (38) must hold. Next, we show that

$$\mathbb{E}[|Y_{N,1}|; P_{Y_{N,1}}] \xrightarrow{N \rightarrow \infty} \mathbb{E}[|Y|; P_Y] < \infty. \quad (40)$$

By (35), (37) and (65) it is implied that

$$\begin{aligned} \mathbb{E}[|Y_{N,1}|; P_{Y_{N,1}}] &= \mathbb{E}[|Y|; P_Y] \\ &+ \frac{1}{\sqrt{N}} \mathbb{E}[|g(\mathbf{X})| \mathbf{h}^T \boldsymbol{\eta}(\mathbf{X}; \boldsymbol{\theta}^*); P_{\mathbf{X}; \boldsymbol{\theta}^*}], \end{aligned}$$

where, by Hölder's inequality

$$\mathbb{E}^2[|g(\mathbf{X})| \mathbf{h}^T \boldsymbol{\eta}(\mathbf{X}; \boldsymbol{\theta}^*); P_{\mathbf{X}; \boldsymbol{\theta}^*}] \leq \mathbb{E}[|g(\mathbf{X})|^2; P_{\mathbf{X}; \boldsymbol{\theta}^*}] \|\mathbf{h}\|_{\mathbf{I}_{\text{FIM}}(\boldsymbol{\theta}^*)}^2.$$

According to (12), (36), (60), assumption A-4, the compactness of Θ and Hölder's inequality [26], there exists a constant $B > 0$ such that $h(\mathbf{x}) \triangleq B u^4(\mathbf{x}) \sum_{r=0}^8 \|\mathbf{x}\|^r \geq |g(\mathbf{x})|^2$ and $\mathbb{E}[h(\mathbf{X}); P_{\mathbf{X}; \boldsymbol{\theta}}]$ is bounded. Hence, by (62) we conclude that the expectation $\mathbb{E}[|g(\mathbf{X})| \mathbf{h}^T \boldsymbol{\eta}(\mathbf{X}; \boldsymbol{\theta}^*); P_{\mathbf{X}; \boldsymbol{\theta}^*}]$ must be bounded, and therefore, the relation (40) must hold.

Finally, by (11), (17), (35)-(38), (40) and Lemma 15.4.1 in [1] we concluded that $\hat{\mathbf{G}}_u(\boldsymbol{\theta}_0) \xrightarrow{N \rightarrow \infty} \mathbf{G}_u(\boldsymbol{\theta}_0)$ \square

D. Auxiliary Lemmas for Proposition 4:

Lemma 1. Assume that conditions A-3-A-5 are satisfied. Under the local alternatives (21), the normalized score function (16) satisfies:

$$\boldsymbol{\eta}_u(\boldsymbol{\theta}) \xrightarrow{N \rightarrow \infty} \mathcal{N}(\mathbf{0}, \mathbf{G}_u(\boldsymbol{\theta}_0)). \quad (41)$$

Proof. Define a triangular array of real random variables obtained from the array in (32):

$$Y_{N,k} \triangleq g(\mathbf{X}_{N,k}; \boldsymbol{\theta}^*), \quad k = 1, \dots, N, \quad N \geq 1, \quad (42)$$

where

$$g(\mathbf{x}) \triangleq u(\mathbf{x}) \mathbf{t}^T \boldsymbol{\psi}_u(\mathbf{x}; \boldsymbol{\theta}) \quad (43)$$

and \mathbf{t} is an arbitrary non-zero vector in \mathbb{R}^m . Since $\mathbf{X}_{N,k}$, $k = 1, \dots, N$ are i.i.d. and the functions $u(\cdot)$ and $\boldsymbol{\psi}_u(\cdot; \cdot)$ are real, then $Y_{N,k}$, $k = 1, \dots, N$ are real and i.i.d.

In the following we show some properties of the statistical moments of $Y_{N,k}$. By Identity 2,

$$\mathbb{E}[Y_{N,k}; P_{Y_{N,1}}] = 0. \quad (44)$$

Moreover, by (11), (42), (43), assumption A-5 and the compactness of Θ there exists a constant $M > 0$ such that

$$\mathbb{E}[Y_{N,k}^2; P_{Y_{N,1}}] = \|\mathbf{t}\|_{\mathbf{G}_u(\boldsymbol{\theta})}^2 \leq M \quad \forall \boldsymbol{\theta} \in \Theta, \quad (45)$$

where under (21) we have that

$$\|\mathbf{t}\|_{\mathbf{G}_u(\boldsymbol{\theta})} \xrightarrow{N \rightarrow \infty} \|\mathbf{t}\|_{\mathbf{G}_u(\boldsymbol{\theta}_0)} > 0. \quad (46)$$

Using (42), (43), (60), assumptions A-3, A-4, the compactness of Θ and Hölder's inequality [26] it can be shown that there exists a constant $M' > 0$, such that

$$\mathbb{E}[|Y_{N,k}|^4; P_{Y_{N,1}}] = \mathbb{E}[g^4(\mathbf{X}); P_{\mathbf{X}; \boldsymbol{\theta}}] \leq M' \quad \forall \boldsymbol{\theta} \in \Theta. \quad (47)$$

Clearly, by (45)

$$s_N^2 \triangleq \sum_{k=1}^N \mathbb{E}[Y_{N,k}^2; P_{Y_{N,1}}] = N \|\mathbf{t}\|_{\mathbf{G}_u(\boldsymbol{\theta})}^2. \quad (48)$$

Therefore, by (46)-(48)

$$\frac{1}{s_N^4} \sum_{k=1}^N \mathbb{E}[|Y_{N,k}|^4; P_{Y_{N,1}}] = \frac{\mathbb{E}[|Y_{N,1}|^4; P_{Y_{N,1}}]}{N \|\mathbf{t}\|_{\mathbf{G}_u(\boldsymbol{\theta})}^4} \xrightarrow{N \rightarrow \infty} 0.$$

Hence, by Lyapounov's central limit theorem [1], (46) and Slutsky's Theorem [26] we conclude that

$$\frac{1}{\sqrt{N}} \sum_{k=1}^N Y_{N,k} \xrightarrow[N \rightarrow \infty]{D} \mathcal{N}(0, \|\mathbf{t}\|_{\mathbf{G}_u(\boldsymbol{\theta}_0)}^2). \quad (49)$$

The relation (41) follows directly from (42), (43), (49) and the Cramér-Wold Device [1]. \square

Lemma 2. Assume that conditions A-3, A-4, A-6, and A-7 are satisfied. Under the local alternatives (21)

$$\hat{\mathbf{F}}_u(\boldsymbol{\theta}^*) \xrightarrow[N \rightarrow \infty]{P} \mathbf{F}_u(\boldsymbol{\theta}_0). \quad (50)$$

Proof. Define a triangular array of real random variables obtained from the array (32):

$$Y_{N,k} \triangleq g(\mathbf{X}_{N,k}; \boldsymbol{\theta}^*), \quad k = 1, \dots, N, \quad N \geq 1, \quad (51)$$

where

$$g(\mathbf{X}; \boldsymbol{\theta}) \triangleq u(\mathbf{X}) [\boldsymbol{\Gamma}_u(\mathbf{x}; \boldsymbol{\theta})]_{l,m} \quad (52)$$

and $\boldsymbol{\Gamma}_u(\cdot; \cdot)$ is defined in (14). Since $\mathbf{X}_{N,k}$, $k = 1, \dots, N$ are i.i.d. and the functions $u(\cdot)$ and $\boldsymbol{\Gamma}_u(\cdot; \cdot)$ are real, then $Y_{N,k}$, $k = 1, \dots, N$ are real and i.i.d. Furthermore, define the random variable

$$Y \triangleq g(\mathbf{X}; \boldsymbol{\theta}_0), \quad (53)$$

where \mathbf{X} has probability distribution $P_{\mathbf{X}; \boldsymbol{\theta}_0}$.

Let $F_{Y_{N,1}}(\cdot)$ and $F_Y(\cdot)$ denote the c.d.f.s of $Y_{N,1}$ and Y , respectively. We show that

$$F_{Y_{N,1}}(y) \xrightarrow[N \rightarrow \infty]{} F_Y(y) \quad \forall y \in C, \quad (54)$$

where $C \subseteq \mathbb{R}$ denotes the set of continuity points of $F_Y(y)$. Let $\zeta_{Y_{N,1}}(t)$ and $\zeta_Y(t)$ denote the characteristic functions of $Y_{N,1}$ and Y , respectively. By (51), (53), assumption A-6 and Identity 3 their difference satisfies:

$$\begin{aligned} |\zeta_{Y_{N,1}}(t) - \zeta_Y(t)| &= \left| \mathbb{E}[e^{ig(\mathbf{X}; \boldsymbol{\theta}^*)t} - e^{ig(\mathbf{X}; \boldsymbol{\theta}_0)t}; P_{\mathbf{X}; \boldsymbol{\theta}_0}] \right| \\ &+ \frac{1}{\sqrt{N}} \mathbb{E} \left[\left| \mathbf{h}^T \boldsymbol{\eta}(\mathbf{X}; \boldsymbol{\theta}^*); P_{\mathbf{X}; \boldsymbol{\theta}^*} \right| \right] \\ &\stackrel{(a)}{\leq} \mathbb{E} \left[\left| e^{ig(\mathbf{X}; \boldsymbol{\theta}^*)t} - e^{ig(\mathbf{X}; \boldsymbol{\theta}_0)t} \right|; P_{\mathbf{X}; \boldsymbol{\theta}_0} \right] \\ &+ \frac{1}{\sqrt{N}} \mathbb{E} \left[\left| \mathbf{h}^T \boldsymbol{\eta}(\mathbf{X}; \boldsymbol{\theta}^*) \right|; P_{\mathbf{X}; \boldsymbol{\theta}^*} \right] \\ &\stackrel{(b)}{\leq} \mathbb{E}[|s(\mathbf{X}; \boldsymbol{\theta}^*, \boldsymbol{\theta}_0, t)|; P_{\mathbf{X}; \boldsymbol{\theta}_0}] \\ &+ \frac{1}{\sqrt{N}} \|\mathbf{h}\|_{\mathbf{I}_{\text{FIM}}(\boldsymbol{\theta}^*)}, \end{aligned} \quad (55)$$

where $\boldsymbol{\eta}(\mathbf{x}; \boldsymbol{\theta}) \triangleq \nabla_{\boldsymbol{\theta}} \log f(\mathbf{x}; \boldsymbol{\theta})$, $\mathbf{I}_{\text{FIM}} \triangleq \mathbb{E}[\boldsymbol{\eta}(\mathbf{x}; \boldsymbol{\theta}) \boldsymbol{\eta}^T(\mathbf{x}; \boldsymbol{\theta})]$ is the Fisher-information matrix [42], and the term $s(\mathbf{x}; \boldsymbol{\theta}, \boldsymbol{\theta}_0, t) \triangleq$

$2 \sin((g(\mathbf{x}; \boldsymbol{\theta}) - g(\mathbf{x}; \boldsymbol{\theta}_0))t)$. We note that (a) follows from the triangle inequality, while (b) follows from Hölder's inequality [26] and the definition of the Fisher-information matrix. Notice that $s(\mathbf{x}; \boldsymbol{\theta}, \boldsymbol{\theta}_0, t)$ is bounded and by (14), (52) and A-3 it is also continuous at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ for any $\mathbf{x} \in \mathcal{X}$. Therefore, since $\boldsymbol{\theta}^* \xrightarrow{N \rightarrow \infty} \boldsymbol{\theta}_0$, by the bounded convergence theorem [26] we conclude that $E[s(\mathbf{X}; \boldsymbol{\theta}^*, \boldsymbol{\theta}_0, t); P_{\mathbf{X}; \boldsymbol{\theta}_0}] \xrightarrow{N \rightarrow \infty} 0$ for any $t \in \mathbb{R}$. Furthermore, by assumption A-7 and (62) the weighted Euclidean norm $\|\mathbf{h}\|_{\mathbf{I}_{\text{FIM}}(\boldsymbol{\theta})}^2$ is bounded over Θ , and therefore, by (55) $\zeta_{Y_N}(t) \xrightarrow{N \rightarrow \infty} \zeta_Y(t)$ for any $t \in \mathbb{R}$. Hence, by Theorem 11.2.2 in [1] (54) must hold.

Next, we show that

$$E[|Y_{N,1}|; P_{Y_{N,1}}] \xrightarrow{N \rightarrow \infty} E[|Y|; P_Y] < \infty. \quad (56)$$

By (51), (53) and (65)

$$\begin{aligned} E[|Y_{N,1}|; P_{Y_{N,1}}] &= E[|g(\mathbf{X}; \boldsymbol{\theta}^*)|; P_{\mathbf{X}; \boldsymbol{\theta}_0}] \\ &+ \frac{E[|g(\mathbf{X}; \boldsymbol{\theta}^*)| \mathbf{h}^T \boldsymbol{\eta}(\mathbf{X}; \boldsymbol{\theta}^*); P_{\mathbf{X}; \boldsymbol{\theta}^*}]}{\sqrt{N}} \end{aligned} \quad (57)$$

According to (14), (52) and assumption A-3 $g(\mathbf{x}, \boldsymbol{\theta})$ is continuous in Θ for any $\mathbf{x} \in \mathcal{X}$. Moreover, by (14), (52), (61), assumptions A-3, A-4, the compactness of Θ and Hölder's inequality [26] there exist positive constants B, C such that:

$$h(\mathbf{x}) \triangleq B \sum_{r=0}^2 \|\mathbf{x}\|^r \geq |g(\mathbf{x}, \boldsymbol{\theta})| \quad (58)$$

and

$$E[h(\mathbf{X}); P_{\mathbf{X}; \boldsymbol{\theta}}] \leq E[h^2(\mathbf{X}); P_{\mathbf{X}; \boldsymbol{\theta}}] \leq C. \quad (59)$$

Therefore, since $\boldsymbol{\theta}^* \xrightarrow{N \rightarrow \infty} \boldsymbol{\theta}_0$, by (53), (58), (59) and the dominated convergence theorem [26] we conclude that the expectation $E[|g(\mathbf{X}; \boldsymbol{\theta}^*)|; P_{\mathbf{X}; \boldsymbol{\theta}_0}] \xrightarrow{N \rightarrow \infty} E[|g(\mathbf{X}; \boldsymbol{\theta}_0)|; P_{\mathbf{X}; \boldsymbol{\theta}_0}] \triangleq E[|Y|; P_Y] < \infty$. We now show that the second summand in (57) converges to zero as $\boldsymbol{\theta}^*$ approaches $\boldsymbol{\theta}_0$. By Hölder's inequality [26], (58) and (59) $\frac{E[|g(\mathbf{x}; \boldsymbol{\theta}^*)| \mathbf{h}^T \boldsymbol{\eta}(\mathbf{x}; \boldsymbol{\theta}^*)]; P_{\mathbf{X}; \boldsymbol{\theta}^*}]^2}{\|\mathbf{h}\|_{\mathbf{I}_{\text{FIM}}(\boldsymbol{\theta}^*)}^2} \leq E[|g(\mathbf{X}; \boldsymbol{\theta}^*)|^2; P_{\mathbf{X}; \boldsymbol{\theta}^*}] \leq E[h^2(\mathbf{X}); P_{\mathbf{X}; \boldsymbol{\theta}^*}] \leq C$. Therefore, by (62) we conclude that the nominator in (57) is bounded, and hence, the right summand in the l.h.s. of (56) approaches zero as $N \rightarrow \infty$.

Finally, by (14), [Fhat], (51)-(54), (56) and Lemma 15.4.1 in [1] we conclude that (50) holds. \square

E. Useful relations and identities:

Relation 1. Assume that the parameter space Θ is compact. Furthermore, assume that $\boldsymbol{\mu}_{\mathbf{x}}^{(u)}(\boldsymbol{\theta})$ and $\boldsymbol{\Sigma}_{\mathbf{x}}^{(u)}(\boldsymbol{\theta})$ are twice continuously differentiable in Θ . Define $d(\mathbf{x}) \triangleq \sum_{r=0}^2 \|\mathbf{x}\|^r$. According to (12), (14) the triangle inequality, the Cauchy-Schwartz inequality and its the matrix extension [43], there exist positive constants B_1 and B_2 , such that for any $(\mathbf{x}, \boldsymbol{\theta}) \in \mathcal{X} \times \Theta$:

$$[\psi_u(\mathbf{x}; \boldsymbol{\theta})]_k \leq B_1 d(\mathbf{x}), \quad \forall k = 1, \dots, m \quad (60)$$

and

$$[\Gamma_u(\mathbf{x}; \boldsymbol{\theta})]_{k,l} \leq B_2 d(\mathbf{x}) \quad \forall k, l = 1, \dots, m. \quad (61)$$

Relation 2. Assume that the parameter space Θ is compact. Furthermore, assume that the Fisher-Information matrix $\mathbf{I}_{\text{FIM}}(\boldsymbol{\theta})$ [42] is continuous in Θ . For any $\mathbf{h} \in \mathbb{R}^m$ there exists a positive constant M such that

$$\|\mathbf{h}\|_{\mathbf{I}_{\text{FIM}}(\boldsymbol{\theta})}^2 \leq M \quad \forall \boldsymbol{\theta} \in \Theta. \quad (62)$$

Identity 1. Assume that the vector function $\psi_u(\mathbf{x}; \boldsymbol{\theta})$ (12) is continuous in Θ for any $\mathbf{x} \in \mathcal{X}$. Furthermore, let $\boldsymbol{\theta} = \boldsymbol{\theta}_0 + \frac{\mathbf{h}}{\sqrt{N}}$. By the mean-value Theorem [44], applied to each entry of $\psi_u(\mathbf{x}; \boldsymbol{\theta})$ and (14), the normalized score function $\eta_u(\boldsymbol{\theta})$ (16) satisfies:

$$\eta_u(\boldsymbol{\theta}_0) = \eta_u(\boldsymbol{\theta}) + \hat{\mathbf{F}}_u(\boldsymbol{\theta}^*) \mathbf{h} \quad (63)$$

where $\hat{\mathbf{F}}_u(\cdot)$ is defined below (25) and $\boldsymbol{\theta}^*$ lies in the line segment connecting $\boldsymbol{\theta}$ and $\boldsymbol{\theta}_0$.

Identity 2. The vector function $\psi_u(\mathbf{X}; \boldsymbol{\theta})$ (12) satisfies:

$$E[u(\mathbf{X}) \psi_u(\mathbf{X}; \boldsymbol{\theta}); P_{\mathbf{X}; \boldsymbol{\theta}}] = \mathbf{0}. \quad (64)$$

[The proof appears in Lemma 6 in [18]]

Identity 3. Let $f(\mathbf{x}; \boldsymbol{\theta}) \triangleq dP_{\mathbf{X}; \boldsymbol{\theta}}(\mathbf{x})/d\rho(\mathbf{x})$ denote the density function of $P_{\mathbf{X}; \boldsymbol{\theta}}$ w.r.t. a σ -finite measure ρ on $\mathcal{S}_{\mathcal{X}}$. Assume that $f(\mathbf{x}; \boldsymbol{\theta})$ is continuous in Θ ρ -a.e. Furthermore, let $\boldsymbol{\theta} = \boldsymbol{\theta}_0 + \frac{\mathbf{h}}{\sqrt{N}}$. By the mean-value Theorem [44]

$$f(\mathbf{x}; \boldsymbol{\theta}) = f(\mathbf{x}; \boldsymbol{\theta}_0) + \frac{\mathbf{h}^T}{\sqrt{N}} \boldsymbol{\eta}(\mathbf{x}; \boldsymbol{\theta}^*) f(\mathbf{x}; \boldsymbol{\theta}^*) \quad \rho\text{-a.e.}, \quad (65)$$

where $\boldsymbol{\eta}(\mathbf{x}; \boldsymbol{\theta}) \triangleq \nabla_{\boldsymbol{\theta}} \log f(\mathbf{x}; \boldsymbol{\theta})$ and $\boldsymbol{\theta}^*$ lies in the line segment connecting $\boldsymbol{\theta}$ and $\boldsymbol{\theta}_0$.

A. REFERENCES

- [1] E. L. Lehmann and J. P. Romano, *Testing Statistical Hypotheses*. Springer Texts in Statistics, 2005.
- [2] C. R. Rao, "Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation," *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 44, no. 1, pp. 50-57, 1948.
- [3] S. M. Silvey, "The Lagrangian multiplier test," *The Annals of Mathematical Statistics*, vol. 30, no. 2, pp. 389-407, 1959.
- [4] L. L. Scharf, *Statistical Signal Processing*. Addison-Wesley, 1991.
- [5] S. M. Kay, *Fundamentals of statistical signal processing: detection theory*. Prentice-Hall, 1998.
- [6] A. Wald, "Tests of statistical hypotheses concerning several parameters when the number of observations is large," *Transactions of the American Mathematical Society*, vol. 54, no. 3, pp. 426-482, 1943.
- [7] H. White, "Maximum likelihood estimation of misspecified models," *Econometrica*, vol. 40, no. 1, pp. 1-25, 1972.
- [8] H. White, *Estimation, inference and specification analysis*. Cambridge university press, 1996.
- [9] R. F. Engle, "Wald, likelihood ratio, and Lagrange multiplier tests in econometrics," *Handbook of econometrics*, vol. 2, pp. 775-826, 1984.
- [10] G. Fiorentini and E. Sentana, "Tests for serial dependence in static, non-Gaussian factor models," *Unobserved components and time series econometrics*, pp. 118-163, Oxford University Press, 2015.
- [11] T. Bollerslev and J. M. Wooldridge, "Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances," *Econometric reviews*, vol. 11, no. 2, pp. 143-172, 1992.
- [12] O. Rousseaux, G. Leus, P. Stoica, "Gaussian maximum-likelihood channel estimation with short training sequences," *IEEE Transactions on Wireless Communications*, vol. 4, no. 6, pp. 2945-2955, 2005.
- [13] E. Weinstein, "Decentralization of the Gaussian maximum likelihood estimator and its applications to passive array processing," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 29, no. 5, pp. 945-951, 1981.
- [14] C. Gourieroux, A. Monfort, and A. Trognon, "Pseudo maximum likelihood methods: Theory," *Econometrica*, vol. 52, no. 3, pp. 681-700, 1984.
- [15] B. Ottersten, M. Viberg, P. Stoica, and A. Nehorai, "Exact and large sample maximum likelihood techniques for parameter estimation and detection," in *Radar Array Processing*, ch. 4, pp. 99-151, Springer-Verlag, 1993.
- [16] H. Krim and M. Viberg, "Two decades of array signal processing research: the parametric approach," *IEEE Signal Processing Magazine*, vol. 13, no. 4, pp. 67-94, 1996.
- [17] K. Todros and A. O. Hero, "Measure-transformed quasi maximum likelihood estimation with application to source localization," *Proceedings of ICASSP 2015*, pp. 3462 - 3466.
- [18] K. Todros and A. O. Hero, "Measure-transformed quasi maximum likelihood estimation," *Accepted for publication in the IEEE Trans. Signal Processing*, 2016. Arxiv version: <https://arxiv.org/pdf/1511.00237.pdf>.
- [19] K. Todros and A. O. Hero, "On measure-transformed canonical correlation analysis," *IEEE Transactions on Signal Processing*, vol. 60, no. 9, pp. 4570-4585, 2012.
- [20] K. Todros and A. O. Hero, "Measure transformed canonical correlation analysis with application to financial data," *Proceedings of SAM 2012*, pp. 361-364, 2012.
- [21] K. Todros and A. O. Hero, "Robust multiple signal classification via probability measure transformation," *IEEE Transactions on Signal Processing*, vol. 63, no. 5, 2015.
- [22] K. Todros and A. O. Hero, "Robust measure transformed MUSIC for DOA estimation," *Proceeding of ICASSP 2014*, pp. 4190-4194, 2014.
- [23] K. Todros and A. O. Hero, "Measure-transformed quasi likelihood ratio test," *Proceedings of ICASSP 2016*, pp. 4259-4263.
- [24] N. Halay, K. Todros and A. O. Hero, "Measure-transformed quasi likelihood ratio test for Bayesian binary hypothesis testing," *Proceedings of SSP 2016*.
- [25] G. B. Folland, *Real Analysis*. John Wiley and Sons, 1984.
- [26] K. B. Athreya and S. N. Lahiri, *Measure theory and probability theory*. Springer-Verlag, 2006.
- [27] F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw and W. A. Stahel, *Robust statistics: the approach based on influence functions*. John Wiley & Sons, 2011.
- [28] S. Kullback and R. A. Leibler, "On information and sufficiency," *The Annals of Mathematical Statistics*, vol. 22, pp. 79-86, 1951.
- [29] P. Schreier and L. L. Scharf, *Statistical signal processing of complex-valued data: the theory of improper and noncircular signals*. p. 39, Cambridge University Press, 2010.
- [30] I. S. Dhillon and J. A. Tropp, "Matrix nearness problems with Bregman divergences," *SIAM Journal on Matrix Analysis and Applications*, vol. 29, no. 4, pp. 1120-1146, 2007.
- [31] R. Horn and C. R. Johnson, *Matrix analysis*. Cambridge University Press, 2012.
- [32] A. Auel and W. Gawronski, "Analytic properties of non-central distributions," *Applied Mathematics and Computation*, vol. 141, pp. 3712, 2003.
- [33] E. Ollila, D. E. Tyler, V. Koivunen and H. V. Poor, "Complex elliptically symmetric distributions: survey, new results and applications," *IEEE Transactions on Signal Processing*, vol. 60, no. 1, pp. 5597-5625, 2012.
- [34] E. Grosicki, A. M. Karim, and H. Yingbo, "A weighted linear prediction method for near-field source localization," *IEEE Transactions on Signal Processing*, vol. 53, no. 10, pp. 3651-3660, 2005.
- [35] M. N. El Korso, R. Boyer, A. Renaux, and S. Marcos, "Conditional and unconditional Cramér-Rao bounds for near-field source localization," *IEEE Transactions on Signal Processing*, vol. 58, no. 5, pp. 2901-2907, 2010.
- [36] A. Swami, "On some parameter estimation problems in alpha-stable processes," *Proceedings of ICASSP-97*, vol. 5, pp. 3541-3544, 1997.
- [37] A. Swami and B. M. Sadler, "TDE, DOA and related parameter estimation problems in impulsive noise," *Proceedings of the IEEE Signal Processing Workshop on Higher-Order Statistics*, 1997, pp. 273-277, 1997.
- [38] A. Swami and B. M. Sadler, "On some detection and estimation problems in heavy-tailed noise," *Signal Processing*, vol. 82, no. 12, pp. 1829-1846, 2002.
- [39] H. B. Mann and A. Wald, "On stochastic limit and order relationships," *Ann. Math. Stat.*, vol. 14, pp. 217-226, 1943.
- [40] R. J. Serfling, *Approximation theorems of mathematical statistics*. John Wiley & Sons, 1980.
- [41] R. Horn and C. R. Johnson, *Matrix analysis*. Cambridge University Press, 2012.
- [42] S. M. Kay, *Fundamentals of statistical signal processing: estimation theory*. Prentice-Hall, 1998.
- [43] X. Yang, "Some trace inequalities for operators," *Journal of the Australian Mathematical Society (Series A)*, vol. 58, no. 3, pp. 281-286, 1995.
- [44] C. H. Edwards, *Advanced calculus of several variables*. Courier Corporation, 2012.